

Conditional assessment of the impact of a Hausman pretest on confidence intervals

Paul Kabaila*, Rheanna Mainzer and Davide Farchione

Department of Mathematics and Statistics, La Trobe University, Australia

JEL classification:

C18

C23

C52

Keywords:

Conditional inference

Coverage probability

Fixed effects model

Hausman specification test

Panel data

Random effects model

* Corresponding author. Department of Mathematics and Statistics, La Trobe University, Victoria 3086, Australia. Tel.: +61 3 9479 2594; fax +61 3 9479 2466. *E-mail address:* P.Kabaila@latrobe.edu.au.

ABSTRACT

We assess the impact of a Hausman pretest, applied to panel data, on a confidence interval for the slope, *conditional* on the observed values of the time-varying covariate. This assessment has the advantages that it (a) relates to the values of this covariate at hand, (b) is valid irrespective of how this covariate is generated, (c) uses finite sample results and (d) results in an assessment that is determined by the values of this covariate and only 2 unknown parameters. Our conditional analysis shows that the confidence interval constructed after a Hausman pretest should not be used.

1 Introduction

For a linear regression model, where the explanatory variables are observed values of random variables, it has long been recognised that, under commonly-occurring circumstances, statistical inference should be carried out conditional on these observed values. Aldrich (2005) describes the seminal contributions of R. A. Fisher (starting in 1922) and M. S. Bartlett to the recognition of this requirement. Adoption of this requirement has the great advantage that the statistical inference is valid irrespective of how the explanatory variables are generated. In the econometric literature, an early recognition of this advantage is provided by Koopmans (1937, pp 29 and 30). A modern description of the justification for this requirement is given in Example 4.3 of Cox (2006). The same justification also applies to linear regression models that include random effects. In particular, this justification applies to the model for longitudinal data that we consider. Statistical inference should be carried out conditional on the observed values of the time-varying covariate. The statistical inference that we consider is a confidence interval for the slope parameter. This confidence interval is assessed by its coverage probability and its scaled expected length, where the scaling is with respect to the expected length of the standard confidence interval with the same minimum coverage probability.

Our aim is to analyze the effect of a Hausman pretest on the coverage probability and scaled expected length of a confidence interval for the slope, *conditional* on the observed values of the time-varying covariate. The four main advantages of this analysis are the following. Firstly, our analysis relates to the values of the time-varying covariate at hand and not to some other values that might have occurred, but are known to not have occurred. Secondly, our analysis has the great advantage that it applies irrespective of the Data Generating Process (DGP) for the time-varying covariate. We do not need to either restrict to some particular DGP, such as a first order autoregression, or concern ourselves with the possible values of the parameters that describe the chosen DGP. Thirdly, our analysis is a finite sample analysis, so that it does not rely on approximations based on large sample results, whose accuracy can be difficult to ascertain in the context of real life sample sizes. Fourthly, as we show, the conditional coverage and scaled expected length of the confidence interval for the slope, constructed after a Hausman pretest, are determined by the time-varying covariate and only 2 unknown parameters: γ which is a scaled version of a non-exogeneity parameter and ν which is the ratio (variance of random effect)

/ (variance of the random error), where $\gamma \in \mathbb{R}$ and $\nu \in (0, \infty)$.

Previous analyses of the effect of a Hausman pretest on either a hypothesis test (Guggenberger, 2010) or a confidence interval (Kabaila, Mainzer and Farchione, 2015) for the slope parameter, have been carried out *unconditionally*. These analyses are restricted to particular DGP's for the time-varying covariate. For example, Kabaila, Mainzer and Farchione (2015) consider two models for the correlation matrix of the time-varying covariate: compound symmetry and first order autoregression. It can be shown that the finite sample *unconditional* coverage and scaled expected length of the confidence interval for the slope, constructed after a Hausman pretest, are determined by 4 known quantities, the unknown parameters γ and ν and also the correlation structure of the time-varying covariate (Kabaila, Mainzer and Farchione, 2015). This additional dependence on this correlation structure is problematic: we are required to assign not only a model for this structure but also plausible ranges of the parameters that describe this model.

By contrast, by Theorems 1 and 4 of the present paper, the finite sample *conditional* coverage and scaled expected length of the confidence interval for the slope, constructed after a Hausman pretest, are determined by the time-varying covariate, either 2 (coverage) or 3 (scaled expected length) known quantities and only two unknown parameters γ and ν . The correlation structure of the time-varying covariate is irrelevant. Let $CP(\gamma, \nu)$ denote the conditional coverage probability of this confidence interval for the slope. Wooldridge (2013) provides a balanced panel data set, *airfare*, used in exercise 14 of Chapter 14. Suppose we are interested in the relationship between *concen* (a measure of market share) and *lfare* (log fare). We use this data to demonstrate the effect of the Hausman pretest on the coverage probability and scaled expected length of a confidence interval for the slope parameter, conditional on the observed values of *concen*, the time-varying covariate.

To assess $CP(\gamma, \nu)$ we could use the *confidence coefficient*. Throughout the paper, we use g and n to denote variables taking values in \mathbb{R} and $(0, \infty)$, respectively. The confidence coefficient is the infimum over both g and n of $CP(g, n)$. Irrespective of the values of γ and ν , $CP(\gamma, \nu)$ is bounded below by the *confidence coefficient*. To find the *confidence coefficient* for the *airline* data we use Theorem 2 of Section 3, which states that $CP(\gamma, \nu)$ is an even function of the unknown parameter γ . We find that the minimum over g of $CP(g, n)$ is an increasing function of n . A graph of this function is shown in Figure 1. All computational results reported in this paper

were found using programs written in R. The *confidence coefficient* of the *airfare* data is approximately 0.19, which is the limit as n approaches 0 of $\min_g CP(g, n)$.

However, this *confidence coefficient* does not utilize the information provided by the data about the unknown parameter ν . If the data strongly contradicts a value of ν near 0 then the *confidence coefficient* is an excessively conservative assessment of $CP(\gamma, \nu)$. An estimate of ν from the *airfare* data is $\hat{\nu} = 12.78$. In Section 3.2 we describe an equi-tailed confidence interval for ν . Theorem 3 gives a pivotal quantity which is key in the construction of this confidence interval. The 98% equi-tailed confidence interval for ν from the *airfare* data is $[11.3976, 14.3829]$. This confidence interval strongly contradicts a value of ν near 0.

We therefore propose the following new assessment of $CP(\gamma, \nu)$. This new assessment is an equi-tailed confidence interval for the minimum over g of $CP(g, \nu)$. For the *airfare* data, the 98% equi-tailed confidence interval for this minimum over g is $[0.8889, 0.9026]$.

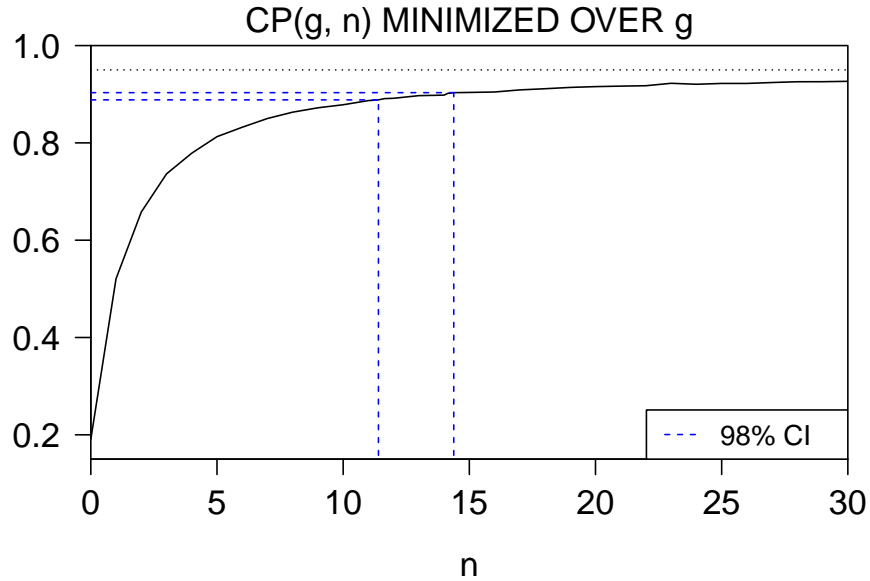


Figure 1: For the *airfare* data, a graph of $\min_g CP(g, n)$, as a function of n . The nominal significance level of the Hausman pretest is $\alpha_H = 0.05$ and the nominal conditional coverage probability of the confidence interval for the slope, constructed after the Hausman pretest, is $1 - \alpha = 0.95$. A 98% confidence interval for ν is given by the points where the dashed vertical lines intersect the horizontal axis and a 98% confidence interval for $\min_g CP(g, \nu)$ is given by the points where the dashed horizontal lines intersect the vertical axis.

2 The model and the practical two-stage procedure

We consider a model for longitudinal data, for which i denotes the individual ($i = 1, \dots, N$) and t denotes the time ($t = 1, \dots, T$). By interpreting i as the cluster index and t as the unit of analysis, our results also apply to the analysis of clustered data. Let y_{it} and x_{it} denote the response variable and the time-varying covariate, respectively, for the i 'th individual at time t . Let $x = (x_{11}, \dots, x_{1T}, \dots, x_{N1}, \dots, x_{NT})$. Our statistical analysis is *conditional* on the observed value of x , so that we treat x as given. Suppose that

$$y_{it} = a + b x_{it} + \xi \bar{x}_i + \eta_i + \varepsilon_{it}. \quad (1)$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$, where $\bar{x}_i = T^{-1} \sum_{t=1}^T x_{it}$. Also suppose that the η_i 's and the ε_{it} 's are independent, with the η_i 's iid $N(0, \sigma_\eta^2)$ and the ε_{it} 's iid $N(0, \sigma_\varepsilon^2)$. The ε_{it} 's and η_i 's are unobserved. This is the “correlated random effects” model described, for example, by Wooldridge (2013). If $\xi = 0$ then the x_{it} 's are exogenous. Thus we call ξ a non-exogeneity parameter.

Suppose that the inference of interest is a confidence interval for the slope parameter b with coverage probability $1 - \alpha$, conditional on x . Assume, for the moment, that σ_ε and σ_η are known. When $\xi = 0$ we can estimate b efficiently from (1) using GLS. Let \hat{b} denote this GLS estimator of b . Also let $\nu = \sigma_\eta^2 / \sigma_\varepsilon^2$ and $z_c = \Phi^{-1}(c)$, where Φ denotes the $N(0, 1)$ cdf. A confidence interval for b that has coverage probability $1 - \alpha$ conditional on x , when $\xi = 0$, is

$$I(\sigma_\varepsilon, \nu) = \left[\hat{b} - z_{1-\alpha/2} \left(\text{Var}(\hat{b} | x) \right)^{1/2}, \hat{b} + z_{1-\alpha/2} \left(\text{Var}(\hat{b} | x) \right)^{1/2} \right],$$

where $\text{Var}(\hat{b} | x)$ denotes the variance of \hat{b} , conditional on x . Adding and subtracting $b \bar{x}_i$ to (1) gives

$$y_{it} = a + b_W (x_{it} - \bar{x}_i) + b_B \bar{x}_i + \eta_i + \varepsilon_{it}, \quad (2)$$

where $b_W = b$ and $b_B = b + \xi$. Note that b_W is the coefficient of a “within” effect and b_B is the coefficient of a “between” effect. Let \tilde{b}_W and \tilde{b}_B denote the GLS estimators of b_W and b_B , respectively, from model (2). A confidence interval for b that has coverage probability $1 - \alpha$ conditional on x , irrespective of the value of ξ , is

$$J(\sigma_\varepsilon) = \left[\tilde{b}_W - z_{1-\alpha/2} \left(\text{Var}(\tilde{b}_W | x) \right)^{1/2}, \tilde{b}_W + z_{1-\alpha/2} \left(\text{Var}(\tilde{b}_W | x) \right)^{1/2} \right],$$

where $\text{Var}(\tilde{b}_W | x)$ denotes the variance of \tilde{b}_W , conditional on x .

As proved in the appendix, \tilde{b}_W and \tilde{b}_B can also be obtained as follows. Averaging (2) over $t = 1, \dots, T$ for each $i = 1, \dots, N$, we obtain the model

$$\bar{y}_i = a + b_B \bar{x}_i + \eta_i + \bar{\varepsilon}_i. \quad (3)$$

where $\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$ and $\bar{\varepsilon}_i = T^{-1} \sum_{t=1}^T \varepsilon_{it}$. The OLS estimator of b_B based on this model is equal to \tilde{b}_B . Subtracting (3) from (2) we obtain the “fixed effects” model

$$y_{it} - \bar{y}_i = b_W (x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i). \quad (4)$$

The OLS estimator of b_W based on this model is equal to \tilde{b}_W .

In practice we do not know whether $\xi = 0$ or not. The usual procedure is to use a Hausman (1978) pretest to test the null hypothesis that $\xi = 0$ ($b_W = b_B$) against the alternative hypothesis that $\xi \neq 0$ ($b_W \neq b_B$). We consider this pretest, based on the test statistic

$$H(\sigma_\varepsilon, \nu) = \frac{(\tilde{b}_W - \tilde{b}_B)^2}{\text{Var}(\tilde{b}_W | x) + \text{Var}(\tilde{b}_B | x)},$$

where $\text{Var}(\tilde{b}_B | x)$ denotes the variance of \tilde{b}_B , conditional on x . This test statistic has a χ_1^2 distribution under the null hypothesis. Suppose that we accept the null hypothesis that $\xi = 0$ if $H(\sigma_\varepsilon, \nu) \leq z_{1-\alpha_H/2}^2$; otherwise we reject the null hypothesis. The level of significance of this test is α_H . We consider the following two-stage procedure. If the null hypothesis is accepted then use the confidence interval $I(\sigma_\varepsilon, \nu)$; otherwise use the confidence interval $J(\sigma_\varepsilon)$. Let $K(\sigma_\varepsilon, \nu)$ denote the confidence interval, with nominal coverage $1 - \alpha$, that results from this two-stage procedure.

Of course, in practice, σ_ε and ν need to be estimated from the data. Let $\hat{\sigma}_\varepsilon$ and $\hat{\nu}$ denote estimators of σ_ε and ν , respectively, described in Section 2.1. Let $H(\hat{\sigma}_\varepsilon, \hat{\nu})$, $I(\hat{\sigma}_\varepsilon, \hat{\nu})$, $J(\hat{\sigma}_\varepsilon)$ and $K(\hat{\sigma}_\varepsilon, \hat{\nu})$ denote the Hausman test statistic, the confidence interval based on \hat{b} , the confidence interval based on \tilde{b}_W and the confidence interval that results from the two-stage procedure, respectively, when σ_ε and ν are replaced by their estimators. Our aim is to assess the coverage probability and expected length properties of $K(\hat{\sigma}_\varepsilon, \hat{\nu})$, conditional on x .

2.1 Estimation of σ_ε and ν

We use the models (3) and (4) to motivate the estimators of σ_ε^2 and ν that we use. Let r_{it} denote the residual for the i 'th individual at the t 'th time when (4)

is estimated by OLS, i.e. let $r_{it} = (y_{it} - \bar{y}_i) - \tilde{b}_W (x_{it} - \bar{x}_i)$ for $i = 1, \dots, N$ and $t = 1, \dots, T$. Then $r_{it} = (b_W - \tilde{b}_W) (x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$. Our estimator

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=1}^T r_{it}^2. \quad (5)$$

is motivated by the approximation $r_{it} \approx \varepsilon_{it} - \bar{\varepsilon}_i$ and the fact that $E(\varepsilon_{it} - \bar{\varepsilon}_i) = 0$ and $\text{Var}(\varepsilon_{it} - \bar{\varepsilon}_i) = N(T-1)\sigma_\varepsilon^2$.

Now we turn to the estimation of σ_η^2 . The most efficient estimator of (a, b_B) , based on the model (3), is the OLS estimator (\tilde{a}, \tilde{b}_B) . Let \tilde{r}_i denote the i 'th residual corresponding to this estimator. In other words, $\tilde{r}_i = \bar{y}_i - (\tilde{a} + \tilde{b}_B \bar{x}_i) = (a - \tilde{a}) + (b_B - \tilde{b}_B) \bar{x}_i + \eta_i + \bar{\varepsilon}_i$. Our estimator

$$\hat{\sigma}_\eta^2 = N^{-1} \sum_{i=1}^N \tilde{r}_i^2 - T^{-1} \hat{\sigma}_\varepsilon^2. \quad (6)$$

is motivated by the approximation $\tilde{r}_i \approx \eta_i + \bar{\varepsilon}_i$ and the fact that $(\eta_i + \bar{\varepsilon}_i)$'s are independent, with $E(\eta_i + \bar{\varepsilon}_i) = 0$ and $\text{Var}(\eta_i + \bar{\varepsilon}_i) = \sigma_\eta^2 + T^{-1} \sigma_\varepsilon^2$. Our estimator of ν is $\hat{\nu} = \hat{\sigma}_\eta^2 / \hat{\sigma}_\varepsilon^2$.

3 The coverage probability of the confidence interval resulting from the two-stage procedure

The coverage probability of the confidence interval $K(\hat{\sigma}_\varepsilon, \hat{\nu})$, conditional on x , is $P(b \in K(\hat{\sigma}_\varepsilon, \hat{\nu}) | x)$. By the law of total probability, $P(b \in K(\hat{\sigma}_\varepsilon, \hat{\nu}) | x)$ is equal to

$$P(b \in I(\hat{\sigma}_\varepsilon, \hat{\nu}), H(\hat{\sigma}_\varepsilon, \hat{\nu}) \leq z_{1-\alpha_H/2}^2 | x) + P(b \in J(\hat{\sigma}_\varepsilon, \hat{\nu}), H(\hat{\sigma}_\varepsilon, \hat{\nu}) > z_{1-\alpha_H/2}^2 | x). \quad (7)$$

Let $\gamma = \xi N^{1/2} / \sigma_\varepsilon$, which is a scaled version of the non-exogeneity parameter ξ . The following two theorems give important properties of this coverage probability. The proofs of these theorems are in the appendix.

Theorem 1. *For the estimators considered in Section 2.1, $P(b \in K(\hat{\sigma}_\varepsilon, \hat{\nu}) | x)$ is determined by x (the time-varying covariate), α_H (the nominal significance level of the Hausman pretest), $1 - \alpha$ (the nominal coverage probability of $K(\hat{\sigma}_\varepsilon, \hat{\nu})$), ν (the ratio $\sigma_\eta^2 / \sigma_\varepsilon^2$) and γ (the scaled non-exogeneity parameter). Given these quantities, the conditional coverage probability does not depend on σ_ε^2 (the variance of the random error) or σ_η^2 (the variance of the random effect).*

Theorem 2. *Suppose that x , α_H , $1 - \alpha$ and ν are fixed. For the estimators considered in Section 2.1, the conditional coverage probability is an even function of γ .*

3.1 Assessment of $CP(\gamma, \nu)$

Suppose that α_H and $1 - \alpha$ are given. Let $CP(\gamma, \nu) = P(b \in K(\hat{\sigma}_\varepsilon, \hat{\nu}) | x)$, the coverage probability of $K(\hat{\sigma}_\varepsilon, \hat{\nu})$, conditional on x . In this section, we ask the question: How do we assess $CP(\gamma, \nu)$? A commonly used assessment of this coverage probability is the *confidence coefficient*, which is the infimum over g and n of $CP(g, n)$. Irrespective of the values of γ and ν , $CP(\gamma, \nu)$ is bounded below by the *confidence coefficient*. We illustrate this assessment using the *airfare* data. For the *airfare* data, the minimum over g of $CP(g, n)$ is an increasing function of n . For this data, the *confidence coefficient* is approximately 0.19.

However, this assessment is excessively conservative, since the data provides information about ν through the estimator $\hat{\nu}$. For any given x , we can assess how much $CP(\gamma, \hat{\nu})$ differs from $CP(\gamma, \nu)$ for a range of values of γ and ν . This has been done for the *airfare* data in Figure 2 by plotting the density function of $CP(\gamma, \hat{\nu})$, estimated by the simulation method described later in Section 5.4. What this figure shows us is that $CP(\gamma, \hat{\nu})$ is unlikely to differ greatly from $CP(\gamma, \nu)$. This suggests that we can provide a useful assessment of $CP(\gamma, \nu)$ by finding an equi-tailed confidence interval for the minimum over g of $CP(g, \nu)$. This confidence interval is constructed from the equi-tailed confidence interval for ν described in the next section. An attractive feature of this assessment is if (hypothetically) the confidence interval for ν is $(0, \infty)$ then this assessment reduces to the confidence coefficient.

3.2 An equi-tailed confidence interval for ν

Using (5) and (6), it can be shown that

$$\frac{\hat{\nu} + T^{-1}}{\nu + T^{-1}} = \frac{\frac{N^{-1} \sum_{i=1}^N \tilde{r}_i^2}{\sigma_\varepsilon^2(\nu + T^{-1})}}{\frac{(N(T-1))^{-1} \sum_{i=1}^N \sum_{t=1}^T r_{it}^2}{\sigma_\varepsilon^2}}. \quad (8)$$

The following theorem allows us to easily compute quantiles of the distribution of $(\hat{\nu} + T^{-1})/(\nu + T^{-1})$ by simulation. This theorem is proved in the appendix.

Theorem 3. *Conditional on x , the distribution of $(\hat{\nu} + T^{-1})/(\nu + T^{-1})$ does not depend on any unknown parameters, i.e. $(\hat{\nu} + T^{-1})/(\nu + T^{-1})$ is a pivotal quantity.*

Define $F_{\bar{\alpha}/2}$ and $F_{1-\bar{\alpha}/2}$ to be the $\bar{\alpha}/2$ and $1 - \bar{\alpha}/2$ quantiles, respectively, of the distribution of the pivotal quantity $(\hat{\nu} + T^{-1})/(\nu + T^{-1})$. Then an equi-tailed

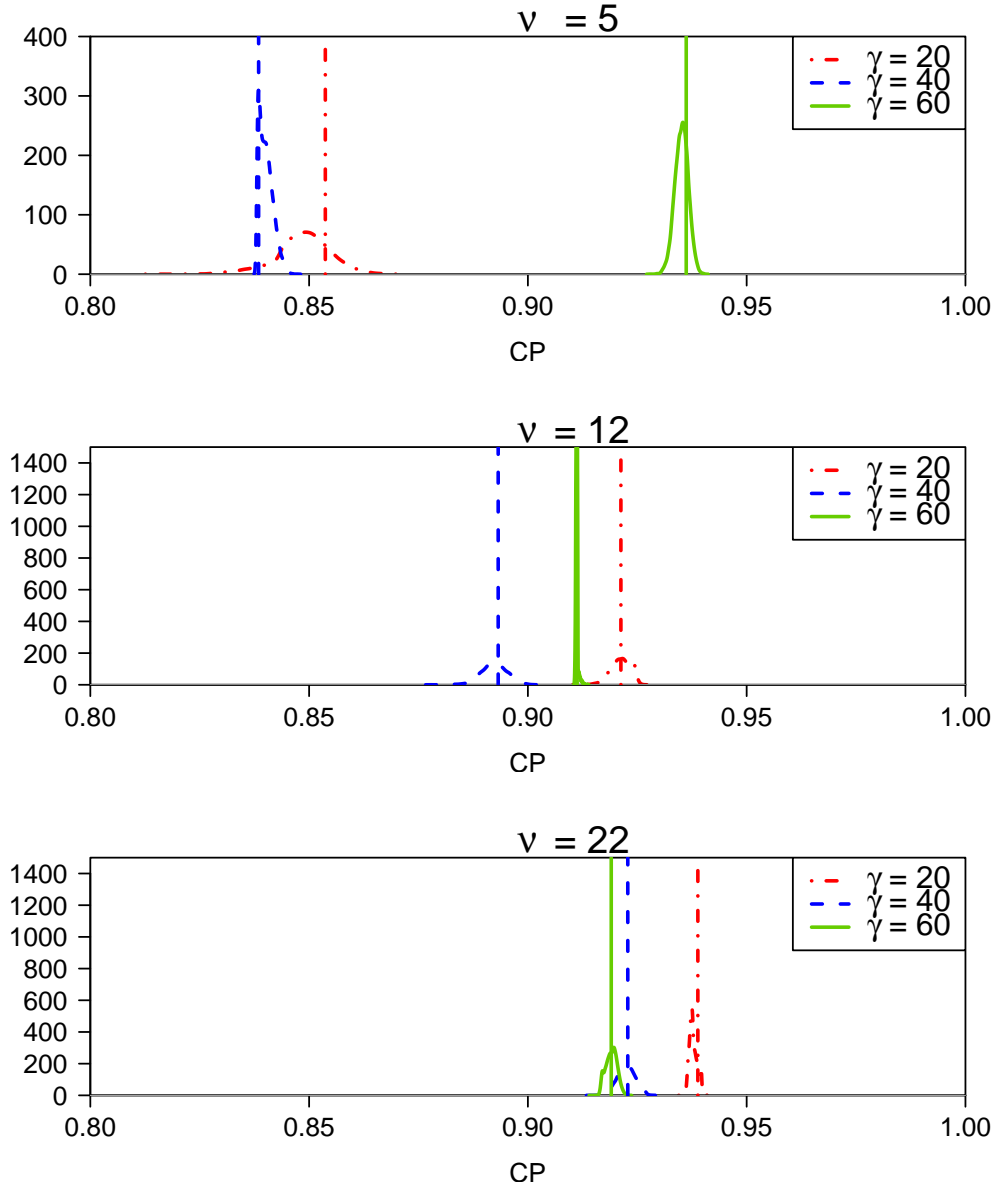


Figure 2: For the *airfare* data, plots of the density function, conditional on x , of $CP(\gamma, \hat{\nu})$, estimated by simulation, for $\nu \in \{5, 12, 22\}$ and $\gamma \in \{20, 40, 60\}$. The vertical lines have horizontal axis intercepts at $CP(\gamma, \nu)$.

confidence interval for ν with coverage probability $1 - \bar{\alpha}$ is

$$\left[\frac{\hat{\nu} + T^{-1}}{F_{1-\bar{\alpha}/2}} - T^{-1}, \frac{\hat{\nu} + T^{-1}}{F_{\bar{\alpha}/2}} - T^{-1} \right]. \quad (9)$$

We find $F_{\bar{\alpha}/2}$ and $F_{1-\bar{\alpha}/2}$ by the simulation method described in detail in Section 5.5. For the *airfare* data described in the introduction, this 98% equi-tailed confidence interval for ν is $[11.3976, 14.3829]$.

3.3 Why don't we use information about γ provided by $\hat{\gamma}$?

An estimator of γ is

$$\hat{\gamma} = \frac{(\tilde{b}_B - \tilde{b}_W) N^{1/2}}{\hat{\sigma}_\varepsilon}. \quad (10)$$

The data provides information about γ through the estimator $\hat{\gamma}$. For any given x , we can assess how much $CP(\hat{\gamma}, \nu)$ differs from $CP(\gamma, \nu)$ for a range of values of γ and ν . This has been done for the *airfare* data in Figure 3 by plotting the probability density function of $CP(\hat{\gamma}, \nu)$, estimated by the simulation method described later in Section 5.4. What this figure shows us is that $CP(\hat{\gamma}, \nu)$ differs greatly from $CP(\gamma, \nu)$. Therefore, we treat the parameter γ differently from the parameter ν .

4 Definition of the conditional scaled expected length

For any $c \in [1/2, 1)$, let

$$J_c(\sigma_\varepsilon) = \left[\tilde{b}_W - \Phi^{-1}((c+1)/2) \left(\text{Var}(\tilde{b}_W | x) \right)^{1/2}, \tilde{b}_W + \Phi^{-1}((c+1)/2) \left(\text{Var}(\tilde{b}_W | x) \right)^{1/2} \right].$$

Note that, for any given c , the confidence interval $J_c(\hat{\sigma}_\varepsilon)$ has coverage probability that does not depend on any unknown parameters. This is the standard confidence interval against which we compare $K(\hat{\sigma}_\varepsilon, \hat{\nu})$, in terms of expected length, conditional on x . As observed in Section 2, this interval may be constructed using the fixed effects model.

The usual definition of conditional scaled expected length of $K(\hat{\sigma}_\varepsilon, \hat{\nu})$ is as follows. Define c_{\min} to be the value of c such that $P(b \in J_c(\hat{\sigma}_\varepsilon)) = \inf_n \min_g CP(g, n)$. Then define this scaled expected length to be the expected length of $K(\hat{\sigma}_\varepsilon, \hat{\nu})$ divided by the expected length of $J_{c_{\min}}(\hat{\sigma}_\varepsilon)$, conditional on x . In other words, we compare the expected length of $K(\hat{\sigma}_\varepsilon, \hat{\nu})$ with the expected length of the standard confidence interval with the same *confidence coefficient*, conditional on x .

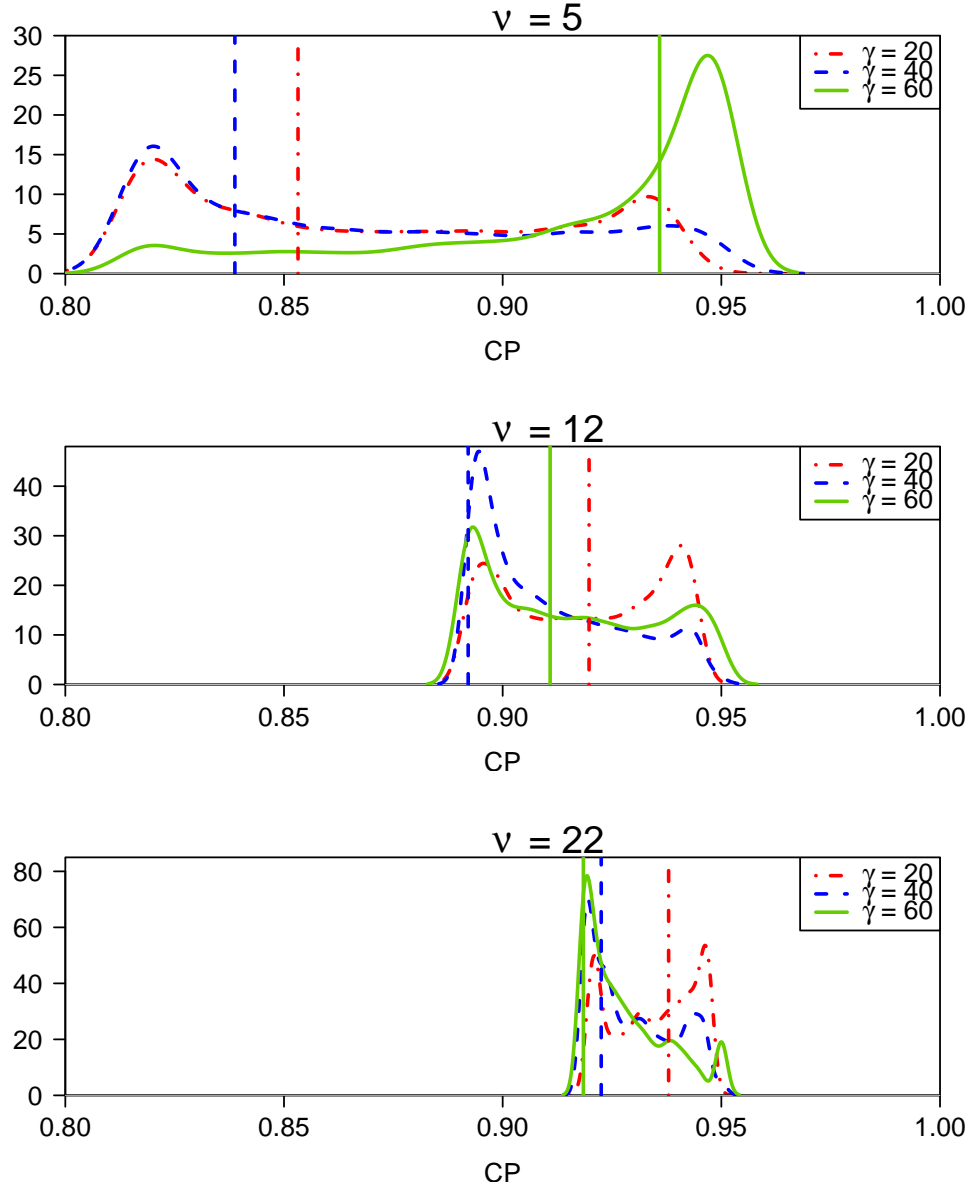


Figure 3: For the *airfare* data, plots of the density function, conditional on x , of $CP(\hat{\gamma}, \nu)$, estimated by simulation, for $\nu \in \{5, 12, 22\}$ and $\gamma \in \{20, 40, 60\}$. The vertical lines have horizontal axis intercepts at $CP(\gamma, \nu)$.

However, as noted in the introduction, the *confidence coefficient* for the airline data is an excessively conservative assessment of $CP(\gamma, \nu)$. We therefore introduce the following alternative definition of conditional scaled expected length. Let $[\nu_\ell, \nu_u]$ be the equi-tailed confidence interval (9) for ν with coverage probability $1 - \bar{\alpha}$. Define c^* to be the value of c such that $P(b \in J_c(\hat{\sigma}_\varepsilon)) = \min_{n \in [\nu_\ell, \nu_u]} \min_g CP(g, n)$. The conditional scaled expected length of $K(\hat{\sigma}_\varepsilon, \hat{\nu})$ is defined to be the expected length of $K(\hat{\sigma}_\varepsilon, \hat{\nu})$ divided by the expected length of $J_{c^*}(\hat{\sigma}_\varepsilon)$, conditional on x . An attractive feature of this definition is that if (hypothetically) the equi-tailed confidence interval for ν with coverage probability $1 - \bar{\alpha}$ is $(0, \infty)$ then c^* is equal to c_{\min} , and this definition of the scaled expected length reduces to the usual definition. Let $\bar{x} = N^{-1} \sum_{i=1}^N \bar{x}_i$, $\text{SSB} = \sum_{i=1}^N (\bar{x}_i - \bar{x})^2$, $\text{SSW} = \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)^2$, $r(x) = \text{SSB}/\text{SSW}$, $q(\hat{\nu}, T) = \hat{\nu} + T^{-1}$, $\hat{w} = q(\hat{\nu}, T) / (q(\hat{\nu}, T) + r(x))$ and define the event $\mathcal{H} = \{H(\hat{\sigma}_\varepsilon, \hat{\nu}) \leq z_{1-\alpha_H/2}^2\}$. For any statement \mathcal{A} , we use the notation

$$\mathcal{I}(\mathcal{A}) = \begin{cases} 1 & \text{if } \mathcal{A} \text{ is true} \\ 0 & \text{if } \mathcal{A} \text{ is false} \end{cases}.$$

The following theorem gives a convenient expression for this alternative definition of the conditional scaled expected length of $K(\hat{\sigma}_\varepsilon, \hat{\nu})$. This theorem is proved in the appendix.

Theorem 4. *For $\gamma \in \mathbb{R}$ and $\nu \in [\nu_\ell, \nu_u]$, the scaled expected length, conditional on x , is equal to*

$$\frac{z_{1-\alpha/2}}{\Phi^{-1}((c^* + 1)/2)} \frac{E((\hat{\sigma}_\varepsilon/\sigma_\varepsilon)(\hat{w}^{1/2} \mathcal{I}(\mathcal{H}) + \mathcal{I}(\mathcal{H}^c)) \mid x)}{E(\hat{\sigma}_\varepsilon/\sigma_\varepsilon)}. \quad (11)$$

The following two theorems give important properties of this scaled expected length. These theorems are proved in the appendix.

Theorem 5. *For the estimators described in Section 2.1 the scaled expected length, conditional on x , is determined by x (the time-varying covariate), α_H (the nominal significance level of the Hausman pretest), $1 - \alpha$ (the nominal coverage probability of $K(\hat{\sigma}_\varepsilon, \hat{\nu})$), $1 - \bar{\alpha}$ (the coverage probability of the confidence interval for ν), ν (the ratio $\sigma_\eta^2/\sigma_\varepsilon^2$) and γ (the scaled non-exogeneity parameter). Given these quantities, the conditional scaled expected length does not depend on σ_ε^2 (the variance of the random error) or σ_η^2 (the variance of the random effect).*

Theorem 6. *Suppose that x , α_H , $1 - \alpha$, $1 - \bar{\alpha}$ and ν are fixed. When σ_ε and σ_μ are replaced by the estimators described in Section 2.1, the conditional scaled expected length is an even function of γ .*

4.1 Numerical illustration for the *airfare* data

Let $SEL(\gamma, \nu)$ denote the conditional scaled expected length of $K(\hat{\sigma}_\varepsilon, \hat{\nu})$. For given x , α_H , $1 - \alpha$ and $1 - \bar{\alpha}$, $SEL(g, \nu)$, minimized over g , is a function of ν . In this section we illustrate the application of $SEL(\gamma, \nu)$ using the *airfare* data. We estimate the scaled expected length using the simulation method described in Section 5.3 for $\alpha_H = 0.05$, $1 - \alpha = 0.95$, $1 - \bar{\alpha} = 0.98$ and $M = 50000$.

As stated in the introduction, the 98% equi-tailed confidence interval for ν is $[11.3976, 14.3829]$. Using Theorem 4, this leads to the 98% equi-tailed confidence interval for $SEL(g, \nu)$, minimized over g , $[1.1012, 1.1244]$. Of course, this confidence interval utilizes the information provided by the data about the unknown parameter ν . Using the usual definition of the scaled expected length that does not utilize this information, we find that $\inf_n \min_g SEL(g, n) = 2.5208$. For both definitions of the scaled expected length, we have nothing to gain by using the confidence interval $K(\hat{\sigma}_\varepsilon, \hat{\nu})$.

5 Simulation methods

In Section 5.1 we give a new theorem that allows us to find a control variate for the estimation by simulation of $CP(\gamma, \nu)$ and $SEL(\gamma, \nu)$ for any given γ and ν . In Sections 5.2 and 5.3 we describe how to estimate $CP(\gamma, \nu)$ and $SEL(\gamma, \nu)$, respectively, by simulation. We also describe how to make use of a control variate for variance reduction. Section 5.4 describes the estimation of the density functions of $CP(\gamma, \hat{\nu})$ and $CP(\hat{\gamma}, \nu)$ shown in Figures 2 and 3, respectively, and Section 5.5 describes the simulation method used to find the quantiles $F_{\bar{\alpha}/2}$ and $F_{1-\bar{\alpha}/2}$ needed for the construction of the equi-tailed confidence interval for ν described in Section 3.2.

5.1 The two-stage procedure when σ_ε and ν are known

In this section we give a new theorem that is used to find a control variate for the estimation by simulation of the coverage probability and scaled expected length of $K(\hat{\sigma}_\varepsilon, \hat{\nu})$. For the rest of this section suppose that σ_ε and ν are known. Define the random variables

$$g_I = \frac{\hat{b} - b}{\left(\text{Var}(\hat{b} | x)\right)^{1/2}}, \quad g_J = \frac{\tilde{b}_W - b}{\left(\text{Var}(\tilde{b}_W | x)\right)^{1/2}}, \quad h = \frac{\tilde{b}_W - \tilde{b}_B}{\left(\text{Var}(\tilde{b}_W | x) + \text{Var}(\tilde{b}_B | x)\right)^{1/2}}.$$

By the law of total probability $P(b \in K(\sigma_\varepsilon, \nu) | x)$, is equal to

$$\begin{aligned} & P(b \in I(\sigma_\varepsilon, \nu), H(\sigma_\varepsilon, \nu) \leq z_{1-\alpha_H/2}^2 | x) + P(b \in J(\sigma_\varepsilon), H(\sigma_\varepsilon, \nu) > z_{1-\alpha_H/2}^2 | x) \\ &= (1 - \alpha) + P(|g_I| \leq z_{1-\alpha/2}, |h| \leq z_{1-\alpha_H/2} | x) - P(|g_J| \leq z_{1-\alpha/2}, |h| \leq z_{1-\alpha_H/2} | x). \end{aligned} \quad (12)$$

The coverage probability of $K(\sigma_\varepsilon, \nu)$ is determined by the distributions of the random vectors (g_I, h) and (g_J, h) . Define $q(\nu, T) = \nu + T^{-1}$ and $w = q(\nu, T)/(q(\nu, T) + r(x))$. Theorem 7 gives the distributions of the random vectors (g_I, h) and (g_J, h) . This theorem is proved in the appendix.

Theorem 7. *Conditional on x , (g_I, h) and (g_J, h) have bivariate normal distributions where*

$$\begin{aligned} E(g_J | x) &= 0, \quad \text{Var}(g_J | x) = 1, \quad E(g_I | x) = \gamma \left(\frac{SSB}{N} \right)^{1/2} \left(\frac{r(x)}{q(\nu, T)(q(\nu, T) + r(x))} \right)^{1/2}, \\ \text{Var}(g_I | x) &= 1, \quad E(h | x) = -\gamma \left(\frac{SSB}{N} \right)^{1/2} \left(\frac{1}{r(x) + q(\nu, T)} \right)^{1/2}, \quad \text{Var}(h | x) = 1, \\ \text{Cov}(g_J, h | x) &= \left(\frac{r(x)}{r(x) + q(\nu, T)} \right)^{1/2}, \quad \text{Cov}(g_I, h | x) = 0. \end{aligned}$$

5.2 Estimation of $CP(\gamma, \nu)$ for given γ and ν by simulation

Consider a grid of γ values. Note that, by the proof of Theorem 1, \widehat{g}_I , \widehat{g}_J and \widehat{h} can be expressed in terms of the ε_{it}^\dagger 's and the η_i^\dagger 's, where the ε_{it}^\dagger 's and η_i^\dagger 's are iid $N(0, 1)$. The simulation method consists of M independent simulation runs. On the k 'th simulation run ($k = 1, \dots, M$) we generate observations of the ε_{it}^\dagger 's and the η_i^\dagger 's and compute \widehat{g}_I , \widehat{g}_J and \widehat{h} for each γ in the grid of values of γ . Let the values of \widehat{g}_I , \widehat{g}_J and \widehat{h} on the k 'th simulation run be denoted by $\widehat{g}_{I,k}$, $\widehat{g}_{J,k}$ and \widehat{h}_k , respectively. Also let $CP = P(b \in K(\widehat{\sigma}_\varepsilon, \widehat{\nu}) | x)$, the coverage probability of $K(\widehat{\sigma}_\varepsilon, \widehat{\nu})$, conditional on x , when σ_ε and ν are unknown. Thus we can define the brute force simulation estimator of CP as

$$\widehat{CP} = \frac{1}{M} \sum_{k=1}^M \left(\mathcal{I}(|\widehat{g}_{I,k}| \leq z_{1-\alpha/2}, |\widehat{h}_k| \leq z_{1-\alpha_H/2}) + \mathcal{I}(|\widehat{g}_{J,k}| \leq z_{1-\alpha/2}, |\widehat{h}_k| > z_{1-\alpha_H/2}) \right).$$

We can also find an estimator of CP that makes use of a control variate for variance reduction. This is done as follows. Let $CPK = P(b \in K(\sigma_\varepsilon, \nu) | x)$, the coverage probability of $K(\sigma_\varepsilon, \nu)$, conditional on x , when σ_ε and ν are known. Let $g_{I,k}$, $g_{J,k}$

and h_k be the values of g_I , g_J and h , respectively, on the k 'th simulation run. An unbiased estimator of CPK is

$$\widehat{\text{CPK}} = \frac{1}{M} \sum_{k=1}^M \left(\mathcal{I}(|g_{I,k}| \leq z_{1-\alpha/2}, |h_k| \leq z_{1-\alpha_H/2}) + \mathcal{I}(|g_{J,k}| \leq z_{1-\alpha/2}, |h_k| > z_{1-\alpha_H/2}) \right).$$

Since we use the same ε_{it}^\dagger 's and η_i^\dagger 's on each simulation run to compute $\widehat{g}_{I,k}$, $\widehat{g}_{J,k}$ and \widehat{h}_k , as we do to compute $g_{I,k}$, $g_{J,k}$ and h_k , we expect that the correlation between $\widehat{\text{CP}}$ and $\widehat{\text{CPK}}$ will be close to 1. We can find CPK exactly using (12) and Theorem 7. Therefore an estimator of CP which makes use of a control variate for variance reduction is

$$\widetilde{\text{CP}} = \widehat{\text{CP}} - (\widehat{\text{CPK}} - \text{CPK}),$$

where $(\widehat{\text{CPK}} - \text{CPK})$ is the control variate which has expected value zero. We estimate the variance of this estimator by noting that it is an average of iid random variables.

An efficiency analysis performed using the airfare data described in the introduction reveals that, if we find the relative efficiency of the control variate estimator to the brute force estimator for, for example, $\nu \in \{11.3976, 14.3829\}$ (the endpoints of the 98% confidence interval for ν) and for every γ in a grid of values, the minimum gain in efficiency is approximately 10.51 and the maximum gain in efficiency is approximately 18.38. In other words, we gain efficiency by using the control variate estimator instead of the brute force estimator.

5.3 Estimation of $SEL(\gamma, \nu)$ for given γ and ν by simulation

Consider a grid of γ values. Let $\text{NUM} = E((\widehat{\sigma}_\varepsilon/\sigma_\varepsilon)(\widehat{w}^{1/2}\mathcal{I}(\mathcal{H}) + \mathcal{I}(\mathcal{H}^c)) | x)$ and $\text{DENOM} = E(\widehat{\sigma}_\varepsilon/\sigma_\varepsilon)$. Then the conditional scaled expected length (11) is equal to

$$\frac{z_{1-\alpha/2}}{\Phi^{-1}((c^* + 1)/2)} \frac{\text{NUM}}{\text{DENOM}}.$$

By the proof of Theorem 5, NUM and DENOM can be expressed in terms of the ε_{it}^\dagger 's and the η_i^\dagger 's, where the ε_{it}^\dagger 's and η_i^\dagger 's are iid $N(0, 1)$. The simulation method consists of M independent simulation runs. Define NUM_k to be the value of $(\widehat{\sigma}_\varepsilon/\sigma_\varepsilon)(\widehat{w}^{1/2}\mathcal{I}(\mathcal{H}) + \mathcal{I}(\mathcal{H}^c))$ and DENOM_k to be the value of $\widehat{\sigma}_\varepsilon/\sigma_\varepsilon$ on the k 'th simulation run ($k = 1, \dots, M$). For one set of M independent simulation runs we do the following. On the k 'th simulation run we generate observations of the ε_{it}^\dagger 's and the η_i^\dagger 's and compute NUM_k . For a second set of M independent simulation runs we do the following. On the k 'th simulation run we generate observations of

the ε_{it}^\dagger 's and compute DENOM_k . Now we can define the brute force simulation estimators

$$\widehat{\text{NUM}} = \frac{1}{M} \sum_{k=1}^M \text{NUM}_k \quad \text{and} \quad \widehat{\text{DENOM}} = \frac{1}{M} \sum_{k=1}^M \text{DENOM}_k.$$

We can also find a control variate estimator of NUM. Define $\mathcal{B} = \{-z_{1-\alpha_H/2} \leq h \leq z_{1-\alpha_H/2}\}$ and $\text{NUMK} = E(w^{1/2}\mathcal{I}(\mathcal{B}) + \mathcal{I}(\mathcal{B}^c) | x)$. Let NUMK_k be the value of $w^{1/2}\mathcal{I}(\mathcal{B}) + \mathcal{I}(\mathcal{B}^c)$ for the k 'th simulation run. Define

$$\widehat{\text{NUMK}} = \frac{1}{M} \sum_{k=1}^M \text{NUMK}_k.$$

Now $\text{NUMK} = w^{1/2}P(\mathcal{B} | x) + 1 - P(\mathcal{B} | x) = 1 + (w^{1/2} - 1)P(\mathcal{B} | x)$, which can be found exactly using Theorem 7. Thus an estimator of NUM which makes use of a control variate for variance reduction is

$$\widetilde{\text{NUM}} = \widehat{\text{NUM}} - (\widehat{\text{NUMK}} - \text{NUMK}),$$

where $(\widehat{\text{NUMK}} - \text{NUMK})$ is the control variate which has expected value zero. Therefore an estimator of the scaled expected length that uses a control variate for variance reduction is

$$\frac{z_{1-\alpha/2}}{\Phi^{-1}((c^* + 1)/2)} \frac{\widetilde{\text{NUM}}}{\widehat{\text{DENOM}}}.$$

We estimate the variance of $\widetilde{\text{NUM}}$ by noting that it is an average of iid random variables.

Similarly to Section 5.2, an efficiency analysis was performed using the airfare data described in the Introduction. This analysis reveals that, if we find the relative efficiency of the control variate estimator to the brute force estimator of NUM for, for example, $\nu \in \{11.3976, 14.3829\}$ (the endpoints of the 98% confidence interval for ν) and for every γ in a grid of values, the minimum gain in efficiency is approximately 0.86 and the maximum gain in efficiency is approximately 10.48. In other words, we gain efficiency by using the control variate estimator instead of the brute force estimator.

5.4 Estimation of the density functions of $CP(\gamma, \hat{\nu})$ and $CP(\hat{\gamma}, \nu)$ for given γ and ν by simulation

Suppose that γ and ν are specified. We begin by describing how to estimate the density function of $CP(\gamma, \hat{\nu})$ by simulation. Note that

$$\frac{\hat{\nu} + T^{-1}}{\nu + T^{-1}} = \left(N^{-1} \sum_{i=1}^N \left(\frac{\tilde{r}_i}{\sigma_\varepsilon} \right)^2 \right) \left((N(T-1))^{-1} \sum_{i=1}^N \sum_{t=1}^T \left(\frac{r_{it}}{\sigma_\varepsilon} \right)^2 \right) \left(\frac{1}{\nu + T^{-1}} \right). \quad (13)$$

In the proof of Theorem 1 we show that $(\tilde{r}_i/\sigma_\varepsilon)$ and $(r_{it}/\sigma_\varepsilon)$ are functions of the ε_{it}^\dagger 's, η_i^\dagger 's, ν and x , where the η_i^\dagger 's and ε_{it}^\dagger 's are iid $N(0, 1)$. Our simulation consist of M independent simulation runs. On the k 'th simulation run ($k = 1, \dots, M$) we generate observations of the η_i^\dagger 's and the ε_{it}^\dagger 's and use these to compute an observation of $(\hat{\nu} + T^{-1}) / (\nu + T^{-1})$ via (13). Thus the M independent simulation runs result in M independent observations of $(\hat{\nu} + T^{-1}) / (\nu + T^{-1})$. We transform these observations using the specified value of ν to obtain M observations of $\hat{\nu}$. Now estimate $CP(\gamma, \nu)$, using the simulation method described in Section 5.2, for an appropriate grid of equally-spaced values of ν . Estimates of $CP(\gamma, \nu)$ for values of ν not on this grid are obtained by linear interpolation. Use this estimate of $CP(\gamma, \nu)$, considered as a function of ν , to transform the M observations of $\hat{\nu}$ into M observations of $CP(\gamma, \hat{\nu})$. We use the **density** function in R to approximate the density function of $CP(\gamma, \hat{\nu})$.

Now we describe how to estimate the density function of $CP(\hat{\gamma}, \nu)$ by simulation. It can be shown that $\hat{\gamma}$, given by (10), is a function of the ε_{it}^\dagger 's, η_i^\dagger 's, γ , ν and x , where the η_i^\dagger 's and ε_{it}^\dagger 's are iid $N(0, 1)$. Our simulation consist of M independent simulation runs. On the k 'th simulation run ($k = 1, \dots, M$) we generate observations of the η_i^\dagger 's and the ε_{it}^\dagger 's and use these to compute an observation of $\hat{\gamma}$. Now estimate $CP(\gamma, \nu)$, using the simulation method described in Section 5.2, for an appropriate grid of equally-spaced values of γ . Estimates of $CP(\gamma, \nu)$ for values of γ not on this grid are obtained by linear interpolation. Use this estimate of $CP(\gamma, \nu)$, considered as a function of γ , to transform the M observations of $\hat{\gamma}$ into M observations of $CP(\hat{\gamma}, \nu)$. We use the **density** function in R to approximate the density function of $CP(\hat{\gamma}, \nu)$.

5.5 Estimation of the quantiles $F_{\bar{\alpha}/2}$ and $F_{1-\bar{\alpha}/2}$ by simulation

To construct the equi-tailed $1 - \bar{\alpha}$ confidence interval (9) for ν , we need to compute $\hat{\nu}$ for our data and find $F_{\bar{\alpha}/2}$ and $F_{1-\bar{\alpha}/2}$, the $\bar{\alpha}/2$ and $1 - \bar{\alpha}/2$ quantiles, respectively, of the distribution of the pivotal quantity $(\hat{\nu} + T^{-1})/(\nu + T^{-1})$. We estimate these quantiles by simulation as follows. Our simulation consist of M independent simulation runs. Since $(\hat{\nu} + T^{-1})/(\nu + T^{-1})$ is a pivotal quantity we specify an arbitrary value of ν , say $\nu = 1$. Similarly to the previous section, on the k 'th simulation run (for $k = 1, \dots, M$) we generate observations of the η_i^\dagger 's and the ε_{it}^\dagger 's and use these to compute an observation of $(\hat{\nu} + T^{-1})/(\nu + T^{-1})$. Thus M independent simulation runs result in M observations of $(\hat{\nu} + T^{-1})/(\nu + T^{-1})$. Arrange these observations in increasing order. Estimate the p 'th quantile of the distribution of $(\hat{\nu} + T^{-1})/(\nu + T^{-1})$ by the r 'th ordered estimate, where r and M are chosen such that $p = r/(M + 1)$. For example, to estimate the 0.01'th quantile of the distribution of $(\hat{\nu} + T^{-1})/(\nu + T^{-1})$ we may choose $r = 100$ and $M = 9999$. A confidence interval for ν is now found using (9), where $F_{1-\bar{\alpha}/2}$ and $F_{\bar{\alpha}/2}$ are replaced by the relevant quantile estimates obtained by simulation.

Appendix

The estimators of b_B and b_W

We can write model (2) in matrix form as $Y = X\beta + u$, where Y , β and u are the column vectors $Y = (y_{11}, \dots, y_{1T}, \dots, y_{N1}, \dots, y_{NT})$, $\beta = (a, b_W, b_B)$ and $u = (\eta_1 + \varepsilon_{11}, \dots, \eta_1 + \varepsilon_{1T}, \dots, \eta_N + \varepsilon_{N1}, \dots, \eta_N + \varepsilon_{NT})$, and X is the matrix

$$X = \begin{bmatrix} 1 & \dots & 1 & \dots & 1 & \dots & 1 \\ x_{11} - \bar{x}_1 & \dots & x_{1T} - \bar{x}_1 & \dots & x_{N1} - \bar{x}_N & \dots & x_{NT} - \bar{x}_N \\ \bar{x}_1 & \dots & \bar{x}_1 & \dots & \bar{x}_N & \dots & \bar{x}_N \end{bmatrix}'.$$

Let I_T denote the $T \times T$ identity matrix and e_T be a T column vector of 1's. The GLS estimator of β is given by $(X'L(\nu)^{-1}X)^{-1}X'L(\nu)^{-1}Y$, where $L(\nu)^{-1}$ is an $NT \times NT$ block diagonal matrix with N identical block diagonal elements each of which is the $T \times T$ matrix $I_T - (\nu/(1 + \nu T))e_T e_T'$. Using this, it can be shown that the GLS

estimators of b_W and b_B are

$$\tilde{b}_W = \frac{\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)}{SSW} = b_W + \frac{\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)}{SSW} \quad (14)$$

$$\tilde{b}_B = \frac{\sum_{i=1}^N (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{SSB} = b_B + \frac{\sum_{i=1}^N (\bar{x}_i - \bar{x})((\eta_i - \bar{\eta}) + (\bar{\varepsilon}_i - \bar{\varepsilon}))}{SSB}. \quad (15)$$

Remember, as defined in Section 4, $SSB = \sum_{i=1}^N (\bar{x}_i - \bar{x})^2$ and $SSW = \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)^2$. The OLS estimator of b_W based on model (4) is (14) and the OLS estimator of b_B based on model (3) is (15). Note that the covariance matrix of the GLS estimator of β is given by $\sigma_\varepsilon^2 (X' L(\nu)^{-1} X)^{-1}$.

The random variables that determine the conditional coverage probability

Let \hat{g}_I , \hat{g}_J and \hat{h} denote the random variables g_I , g_J and h introduced in Section 5.1, when σ_ε and ν have been replaced by their estimators. By Section 3, $P(b \in K(\hat{\sigma}_\varepsilon, \hat{\nu}) | x)$ is equal to

$$\begin{aligned} & P(b \in I(\hat{\sigma}_\varepsilon, \hat{\nu}), H(\hat{\sigma}_\varepsilon, \hat{\nu}) \leq z_{1-\alpha_H/2}^2 | x) + P(b \in J(\hat{\sigma}_\varepsilon), H(\hat{\sigma}_\varepsilon, \hat{\nu}) > z_{1-\alpha_H/2}^2 | x) \\ & = P(|\hat{g}_I| \leq z_{1-\alpha/2}, |\hat{h}| \leq z_{1-\alpha_H/2}^2 | x) + P(|\hat{g}_J| \leq z_{1-\alpha/2}, |\hat{h}| > z_{1-\alpha_H/2}^2 | x). \end{aligned}$$

The following lemma gives expressions for \hat{g}_I , \hat{g}_J and \hat{h} that are used in the proofs of the theorems.

Lemma 1.

$$\begin{aligned} \hat{g}_I &= \frac{q(\hat{\nu}, T) \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(\varepsilon_{it} - \bar{\varepsilon}_i) + \sum_{i=1}^N (\bar{x}_i - \bar{x})((\eta_i - \bar{\eta}) + (\bar{\varepsilon}_i - \bar{\varepsilon}))}{\hat{\sigma}_\varepsilon (q(\hat{\nu}, T) SSW (q(\hat{\nu}, T) + r(x)))^{1/2}} \\ &\quad + \gamma \left(\frac{r(x)}{q(\hat{\nu}, T) (q(\hat{\nu}, T) + r(x))} \right)^{1/2} \left(\frac{SSB}{N} \right)^{1/2} \left(\frac{\sigma_\varepsilon}{\hat{\sigma}_\varepsilon} \right) \\ \hat{g}_J &= \frac{\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(\varepsilon_{it} - \bar{\varepsilon}_i)}{\hat{\sigma}_\varepsilon (SSW)^{1/2}} \\ \hat{h} &= \frac{r(x) \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(\varepsilon_{it} - \bar{\varepsilon}_i) - \sum_{i=1}^N (\bar{x}_i - \bar{x})((\eta_i - \bar{\eta}) + (\bar{\varepsilon}_i - \bar{\varepsilon}))}{\hat{\sigma}_\varepsilon (SSB (q(\hat{\nu}, T) + r(x)))^{1/2}} \\ &\quad - \gamma \left(\frac{1}{r(x) + q(\hat{\nu}, T)} \right)^{1/2} \left(\frac{SSB}{N} \right)^{1/2} \left(\frac{\sigma_\varepsilon}{\hat{\sigma}_\varepsilon} \right). \end{aligned}$$

Proof. It can be shown that

$$\tilde{b}_W \sim N\left(b_W, \frac{\sigma_\varepsilon^2}{\text{SSW}}\right) \quad \text{and} \quad \tilde{b}_B \sim N\left(b_B, \frac{\sigma_\varepsilon^2(\nu + T^{-1})}{\text{SSB}}\right). \quad (16)$$

Thus $\text{Var}(\tilde{b}_W | x) = \sigma_\varepsilon^2 / \text{SSW}$ and $\text{Var}(\tilde{b}_B | x) = (\sigma_\varepsilon^2(\nu + T^{-1})) / \text{SSB}$. Lemma 1 follows from this, equations (14) and (15), and Maddala's (1971) equality $\hat{b} = \hat{w}\tilde{b}_W + (1 - \hat{w})\tilde{b}_B$. \square

Proof of Theorem 1

Let $\varepsilon_{it}^\dagger = \varepsilon_{it} / \sigma_\varepsilon$, $\bar{\varepsilon}_i^\dagger = \bar{\varepsilon}_i / \sigma_\varepsilon$, $\bar{\varepsilon}^\dagger = \bar{\varepsilon} / \sigma_\varepsilon$, $\eta_i^\dagger = \eta_i / \sigma_\eta$ and $\bar{\eta}^\dagger = \bar{\eta} / \sigma_\eta$. The η_i^\dagger 's and ε_{it}^\dagger 's are iid $N(0, 1)$. It can be shown that

$$\begin{aligned} \frac{r_{it}}{\sigma_\varepsilon} &= -\frac{(x_{it} - \bar{x}_i) \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(\varepsilon_{it}^\dagger - \bar{\varepsilon}_i^\dagger)}{\text{SSW}} + (\varepsilon_{it}^\dagger - \bar{\varepsilon}_i^\dagger) \\ \frac{\tilde{r}_i}{\sigma_\varepsilon} &= -\left(\frac{(\nu^{1/2} \bar{\eta}^\dagger + \bar{\varepsilon}^\dagger) \sum_{i=1}^N \bar{x}_i^2 - \bar{x} \sum_{i=1}^N \bar{x}_i (\nu^{1/2} \eta_i^\dagger + \bar{\varepsilon}_i^\dagger)}{\text{SSB}} \right) \\ &\quad - \left(\frac{\sum_{i=1}^N (\bar{x}_i - \bar{x}) (\nu^{1/2} (\eta_i^\dagger - \bar{\eta}^\dagger) + (\bar{\varepsilon}_i^\dagger - \bar{\varepsilon}^\dagger))}{\text{SSB}} \right) \bar{x}_i + \nu^{1/2} \eta_i^\dagger + \bar{\varepsilon}_i^\dagger. \end{aligned}$$

Thus $\hat{\sigma}_\varepsilon / \sigma_\varepsilon$ and $\hat{\sigma}_\eta / \sigma_\eta$ can be expressed in terms of the ε_{it}^\dagger 's, η_i^\dagger 's, ν and x . Hence $\hat{\nu} = \hat{\sigma}_\eta / \hat{\sigma}_\varepsilon$ can be expressed in terms of the ε_{it}^\dagger 's, η_i^\dagger 's, ν and x . Now, using Lemma 1, divide the numerator and denominator of \hat{g}_I , \hat{g}_J and \hat{h} by σ_ε . It follows that \hat{g}_I , \hat{g}_J and \hat{h} can be expressed in terms of ε_{it}^\dagger 's, η_i^\dagger 's, ν , γ and x .

Proof of Theorem 2

Define $\varepsilon = (\varepsilon_{11}, \dots, \varepsilon_{1J}, \dots, \varepsilon_{N1}, \dots, \varepsilon_{NJ})$ and $\eta = (\eta_1, \dots, \eta_N)$. Introduce the notation $\hat{g}_I = \hat{g}_I(x, \varepsilon, \eta, \gamma)$, $\hat{g}_J = \hat{g}_J(x, \varepsilon)$ and $\hat{h} = \hat{h}(x, \varepsilon, \eta, \gamma)$ to show the dependence of \hat{g}_I , \hat{g}_J and \hat{h} on x, ε, η and γ . Note that $\hat{g}_J(x, \varepsilon)$ is not a function of γ . We have $\{|\hat{g}_I| \leq z_{1-\alpha/2}\} = \{-z_{1-\alpha/2} \leq -\hat{g}_I \leq z_{1-\alpha/2}\}$, $\{|\hat{h}| \leq z_{1-\alpha_H/2}\} = \{-z_{1-\alpha_H/2} \leq -\hat{h} \leq z_{1-\alpha_H/2}\}$ and $\{|\hat{h}| > z_{1-\alpha_H/2}\} = \{-\hat{h} < z_{1-\alpha_H/2}\} \cup \{\hat{h} > z_{1-\alpha_H/2}\}$. Thus for the coverage probability to be an even function of γ , it is sufficient to prove that (a) the distribution of $(\hat{g}_I(x, \varepsilon, \eta, d), \hat{h}(x, \varepsilon, \eta, d))$ is the same as the distribution of $(-\hat{g}_I(x, \varepsilon, \eta, -d), -\hat{h}(x, \varepsilon, \eta, -d))$ and (b) the distribution of $(\hat{g}_J(x, \varepsilon), \hat{h}(x, \varepsilon, \eta, d))$ is the same as the distribution of $(-\hat{g}_J(x, \varepsilon), -\hat{h}(x, \varepsilon, \eta, -d))$. For the sake of brevity, we only give the proof of (a). The proof of (b) is similar.

Let $\varepsilon_{it}^* = -\varepsilon_{it}$ for $i = 1, \dots, N$ and $t = 1, \dots, T$, and $\varepsilon^* = (-\varepsilon_{11}, \dots, -\varepsilon_{1J}, \dots, -\varepsilon_{N1}, \dots, -\varepsilon_{NJ})$. Also let $\eta_i^* = -\eta_i$ for $i = 1, \dots, N$ and $\eta^* = (-\eta_1, \dots, -\eta_N)$. Since the ε_{it} 's and η_i 's are independent, the ε_{it} 's are iid $N(0, \sigma_\varepsilon^2)$ and the η_i 's are iid $N(0, \sigma_\eta^2)$, it follows that the ε_{it}^* 's and η_i^* 's are independent, the ε_{it}^* 's are iid $N(0, \sigma_\varepsilon^2)$ and the η_i^* 's are iid $N(0, \sigma_\eta^2)$. Therefore $(\widehat{g}_I(x, \varepsilon^*, \eta^*, d), \widehat{h}(x, \varepsilon^*, \eta^*, d))$ has the same distribution as $(\widehat{g}_I(x, \varepsilon, \eta, d), \widehat{h}(x, \varepsilon, \eta, d))$. Using Lemma 1, it can be shown that $(-\widehat{g}_I(x, \varepsilon, \eta, d), -\widehat{h}(x, \varepsilon, \eta, d))$ has the same distribution as $(\widehat{g}_I(x, \varepsilon^*, \eta^*, -d), \widehat{h}(x, \varepsilon^*, \eta^*, -d))$. It follows that the distribution of $(\widehat{g}_I(x, \varepsilon, \eta, d), \widehat{h}(x, \varepsilon, \eta, d))$ is the same as the distribution of $(-\widehat{g}_I(x, \varepsilon, \eta, -d), -\widehat{h}(x, \varepsilon, \eta, -d))$.

Proof of Theorem 3

The following lemma defines new quantities whose joint distribution does not depend on any unknown parameters.

Lemma 2. *For $i = 1, \dots, N$ and $t = 1, \dots, T$, define*

$$\vartheta_{it}^\dagger = \frac{\varepsilon_{it} - \bar{\varepsilon}_i}{\sigma_\varepsilon} \quad \text{and} \quad \varphi_i^\dagger = \frac{\eta_i + \bar{\varepsilon}_i}{(\sigma_\varepsilon^2 (\nu + J^{-1}))^{1/2}}.$$

Then the random vector $(\varphi_1^\dagger, \dots, \varphi_N^\dagger, \vartheta_{11}^\dagger, \dots, \vartheta_{1T}^\dagger, \dots, \vartheta_{N1}^\dagger, \dots, \vartheta_{NT}^\dagger)$ has a multivariate normal distribution with mean vector 0 and covariance matrix which does not depend on any unknown parameters.

Proof. We assume that the ε_{it} 's and η_i 's are independent, that the ε_{it} 's are iid $N(0, \sigma_\varepsilon^2)$ and that the η_i 's are iid $N(0, \sigma_\eta^2)$. Thus $E(\varepsilon_{it} - \bar{\varepsilon}_i) = 0$, $\text{Cov}(\varepsilon_{it}, \bar{\varepsilon}_i) = T^{-1} \sigma_\varepsilon^2$, $\text{Var}(\varepsilon_{it} - \bar{\varepsilon}_i) = \sigma_\varepsilon^2 (1 - T^{-1})$, $E(\eta_i + \bar{\varepsilon}_i) = 0$, $\text{Cov}(\eta_i, \bar{\varepsilon}_i) = 0$ and $\text{Var}(\eta_i + \bar{\varepsilon}_i) = \sigma_\varepsilon^2 (\nu + T^{-1})$. It follows that $E(\vartheta_{it}^\dagger) = 0$, $E(\varphi_i^\dagger) = 0$, $\text{Var}(\vartheta_{it}^\dagger) = 1 - T^{-1}$ and $\text{Var}(\varphi_i^\dagger) = 1$. Also, for $i \neq k$ and $t \neq s$, we have $\text{Cov}(\vartheta_{it}^\dagger, \vartheta_{is}^\dagger) = T^{-1}$, $\text{Cov}(\vartheta_{it}^\dagger, \vartheta_{kt}^\dagger) = 0$, $\text{Cov}(\varphi_i^\dagger, \varphi_k^\dagger) = 0$, $\text{Cov}(\varphi_i^\dagger, \vartheta_{it}^\dagger) = 0$ and $\text{Cov}(\varphi_k^\dagger, \vartheta_{it}^\dagger) = 0$. \square

The denominator of (8) can be written as

$$\frac{\sum_{i=1}^N \sum_{t=1}^T (\vartheta_{it}^\dagger)^2}{N(T-1)} - \frac{\left(\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i) \vartheta_{it}^\dagger \right)^2}{N(T-1) \text{SSW}},$$

which is in terms of the ϑ_{it}^\dagger 's and x . In a similar manner, the numerator of (8) can be written in terms of the ϑ_{it}^\dagger 's, φ_i^\dagger 's and x . It follows from Lemma 2 that the distribution of (8) does not depend on any unknown parameters.

Proof of Theorem 4

It can be shown that the length of $I(\widehat{\sigma}_\varepsilon, \widehat{\nu})$ is $2 z_{1-\alpha/2} \widehat{\sigma}_\varepsilon \text{SSW}^{-1/2} \widehat{w}^{1/2}$ and that the length of $J(\widehat{\sigma}_\varepsilon)$ is $2 z_{1-\alpha/2} \widehat{\sigma}_\varepsilon \text{SSW}^{-1/2}$. Hence the length of $K(\widehat{\sigma}_\varepsilon, \widehat{\nu})$ is equal to $2 z_{1-\alpha/2} \widehat{\sigma}_\varepsilon \text{SSW}^{-1/2} (\widehat{w}^{1/2} \mathcal{I}(\mathcal{H}) + \mathcal{I}(\mathcal{H}^c))$. Also, the length of $J_{c^*}(\widehat{\sigma}_\varepsilon)$ is equal to $2 \Phi^{-1}((c^* + 1)/2) \widehat{\sigma}_\varepsilon \text{SSW}^{-1/2}$. Thus the scaled expected length is equal to

$$\frac{E(2 z_{1-\alpha/2} \widehat{\sigma}_\varepsilon \text{SSW}^{-1/2} (\widehat{w}^{1/2} \mathcal{I}(\mathcal{H}) + \mathcal{I}(\mathcal{H}^c)) \mid x)}{E(2 \Phi^{-1}((c^* + 1)/2) \widehat{\sigma}_\varepsilon \text{SSW}^{-1/2})},$$

which can be simplified to obtain (11).

Proof of Theorem 5

Consider the conditional scaled expected length expression given by (11). In the proof of Theorem 1 we have shown that $(\widehat{\sigma}_\varepsilon/\sigma_\varepsilon)$ is a function of the ε_{it}^\dagger 's and x . Next, c^* depends on $1 - \bar{\alpha}$, and $\mathcal{I}(\mathcal{H})$ and $\mathcal{I}(\mathcal{H}^c)$ depend on α_H and \widehat{h} . It follows from the proof of Theorem 1 that \widehat{h} is a function of the ε_{it}^\dagger 's, η_i^\dagger 's, ν , γ and x . We have also shown in the proof of Theorem 1 that $\widehat{\nu}$ is a function of the ε_{it}^\dagger 's, η_i^\dagger 's, ν and x , which implies that $\widehat{w} = q(\widehat{\nu}, T)/(q(\widehat{\nu}, T) + r(x))$ is a function of the ε_{it}^\dagger 's, η_i^\dagger 's, ν and x .

Proof of Theorem 6

The conditional scaled expected length (11) depends on γ only through $\mathcal{I}(\mathcal{H})$ and $\mathcal{I}(\mathcal{H}^c)$. Similarly to the proof of Theorem 2, we write $\widehat{h} = \widehat{h}(x, \varepsilon, \eta, \gamma)$ to emphasize the dependence of \widehat{h} on x, ε, η and γ . In the proof of Theorem 2 we have shown that the distribution of $\widehat{h}(x, \varepsilon, \eta, d)$ is the same as the distribution of $-\widehat{h}(x, \varepsilon, \eta, -d)$. It follows that both $\mathcal{I}(\mathcal{H})$ and $\mathcal{I}(\mathcal{H}^c)$ are even functions of γ . Therefore the conditional scaled expected length is an even function of γ .

Proof of Theorem 7

It follows from (16), Maddala's (1971) equality and the fact that $\text{Cov}(\tilde{b}_W, \tilde{b}_B | x) = 0$, that

$$\begin{aligned}\hat{b} - b &\sim N\left(\frac{\xi r(x)}{q(\nu, T) + r(x)}, \frac{\sigma_\varepsilon^2}{\text{SSW}} \frac{q(\nu, T)}{q(\nu, T) + r(x)}\right), \\ \tilde{b}_W - b &\sim N\left(0, \frac{\sigma_\varepsilon^2}{\text{SSW}}\right), \\ \tilde{b}_W - \tilde{b}_B &\sim N\left(-\xi, \frac{\sigma_\varepsilon^2}{\text{SSW}} \left(1 + \frac{q(\nu, T)}{r(x)}\right)\right).\end{aligned}$$

Now use the definitions of g_I , g_J and h and find the relevant expected values, variances and covariances.

References

- Aldrich, J., 2005. Fisher and regression. *Statistical Science*, 20, 401–417.
- Cox, D.R., 2006. *Principles of Statistical Inference*. Cambridge University Press: Cambridge.
- Guggenberger, P., 2010. The impact of a Hausman pretest on the size of a hypothesis test: the panel data case. *Journal of Econometrics*, 156, 337–343.
- Hausman, J. A., 1978. Specification test in econometrics. *Econometrica*, 46, 1251–1271.
- Kabaila, P., Mainzer, R., Farchione, D., 2015. The impact of a Hausman pretest, applied to panel data, on the coverage probability of confidence intervals. *Economics Letters*, 131, 12–15.
- Koopmans, T.C., 1937. *Linear Regression Analysis of Economic Time Series*. Bohn, Haarlem, Netherlands.
- Maddala, G. S., 1971. The use of variance components models in pooling cross section and time series data. *Econometrica*, 39, 341 – 358.
- Wooldridge, J. M., 2013. *Introductory econometrics: a modern approach*. 5th edition. South-Western: Ohio.